

具有智能交互学习能力的机械臂写字系统*

杨明浩^{1,3}, 张珂^{1,2}, 赵博程^{1,4}, 朱庆杰^{1,2}, 潘航^{1,2}, 那桑若阳¹, 湛永松², 陶建华^{1,3,4}



¹(模式识别国家重点实验室(中国科学院 自动化研究所),北京 100190)

²(桂林电子科技大学,广西 桂林 541004)

³(中国科学院 自动化研究所 脑科学与智能技术卓越创新中心,北京 100190)

⁴(中国科学院大学,北京 100190)

通讯作者: 杨明浩, E-mail: mhyang@nlpr.ia.ac.cn

摘要: 基于 Uarm 机械臂构建了一个学习人类写字顺序的机械臂智能写字系统,该系统首先具有对陌生汉字的自动笔画拆分和书写能力,然后基于语音对话和图像分析技术,能够根据用户教授的笔画和笔顺来学习汉字的正确书写方式.首先,系统根据输入的语音信息以及摄像头观察到的文字的图像信息,获得用户想要写的关键字及对话意图;然后通过对摄像头看到的图像信息进行分析,对检测到的汉字进行自动笔画拆分和笔顺提取,对于正在教授中的字,跟踪笔迹顺序,学习汉字笔顺的正确写法.通过对话管理,机械臂会以对话的形式进行书写反馈并与用户交互,学习人类书写顺序并实现正确书写.通过实验分析及测试者主观评测,该系统取得了不错的评价.

关键词: 写字机器人;智能机械臂;交互学习;智能系统

中文引用格式: 杨明浩,张珂,赵博程,朱庆杰,潘航,那桑若阳,湛永松,陶建华.具有智能交互学习能力的机械臂写字系统.软件学报,2018,29(Suppl.(2)):54-61. <http://www.jos.org.cn/1000-9825/18017.htm>

英文引用格式: Yang MH, Zhang K, Zhao BC, Zhu QJ, Pan H, Na SRY, Zhan YS, Tao JH. Robotic writing system with intelligent interactive learning ability. Ruan Jian Xue Bao/Journal of Software, 2018,29(Suppl.(2)):54-61 (in Chinese). <http://www.jos.org.cn/1000-9825/18017.htm>

Robotic Writing System with Intelligent Interactive Learning Ability

YANG Ming-Hao^{1,3}, ZHANG Ke^{1,2}, ZHAO Bo-Cheng^{1,4}, ZHU Qing-Jie^{1,2}, PAN Hang^{1,2}, NA Shen-Ruo-Yang¹, ZHAN Yong-Song², TAO Jian-Hua^{1,3,4}

¹(National Laboratory of Pattern Recognition (Institute of Automation, The Chinese Academy of Sciences), Beijing 100190, China)

²(Guilin University of Electronic Technology, Guilin 541004, China)

³(Research Center for Brain-Inspired Intelligence (Institute of Automation, The Chinese Academy of Sciences), Beijing 100190, China)

⁴(University of Chinese Academy of Sciences, Beijing 100190, China)

基金项目: 国家重点研发项目(2017YFB1002804); 国家自然科学基金(61873269, 61332017); 广西壮族自治区自然科学基金(2017GXNSFAA198226); 广西科技计划(桂科 AB17195053); 广西高校图像图形智能处理重点实验室课题(GIIP201602); 广西可信软件重点实验室研究课题(KX201601); 广西云计算与大数据协同创新中心、广西高校云计算与复杂系统重点实验室资助项目(YD16E11); 广西密码学与信息安全重点实验室研究课题(CIS201602); 桂林电子科技大学研究生教育创新计划(2017YJXC55); 2017年广西壮族自治区自然科学基金课题(2017JJA160182); 2015年广西科技攻关课题(1598018-6)

Foundation item: National Key Research & Development Plan of China (2017YFB1002804); National Nature Science Foundation of China (61873269, 61332017); Guangxi Zhuang Autonomous Region Natural Science Foundation of China (2017GXNSFAA198226); Guangxi Science and Technology Project (桂科 AB17195053); Guangxi Colleges and Universities Key Laboratory of Intelligent Processing of Computer Images and Graphics (GIIP201602); Guangxi Key Laboratory of Trusted Software (KX201601); Guangxi Cooperative Innovation Center of Cloud Computing and Big Data and Guangxi Colleges and Universities Key Laboratory of Cloud Computing and Complex Systems (YD16E11); Guangxi Key Laboratory of Cryptography and Information Security (CIS201602); Innovation Project of GUET Graduate Education (2017YJXC55); Guangxi Zhuang Autonomous Region Natural Science Foundation of China, Project 2017 (2017JJA160182); Guangxi Science and Technology Research Project (1598018-6)

收稿时间: 2017-06-01; 采用时间: 2018-11-08

Abstract: In this study, a robotic intelligence writing system is built based on the Uarm to learn Chinese character strokes. This system can finish automatically strokes splitting and writing of unfamiliar character. Besides, based on the dialogue technology and image processing technology, the system can learn the correct strokes from human. Firstly, the system gets the keyword which user want to write and user intention according to the input voice information and the word image information from camera. Then it analyzes the word image and splitting and extracting the strokes if the keyword is detected. If the word is being taught by human, the system would record the strokes order and learn the correct way to write the character. Through the dialogue management, the Uarm can interact with human through writing and dialogue, learn from human, and write the characters correctly. According to the experimental analysis and subjective evaluation of the test, the system has been well recognized.

Key words: writing robot; intelligence manipulator; interactive learning; intelligence system

1 引言

随着语音识别、图像处理及自然语言理解等技术的进步,机器人正在由预先编程向智能化发展,这些智能技术赋予了机器人更多的交互手段^[1],使得机器人不仅可以通过交互完成既定的工作,还可以在交互过程中学习更多新的知识.让机器人具有交互学习能力在很多方面具有重要体现.例如,在教育方面,可以让机器人来教人学英语、学写字等,突破传统由人来教的局限性,也增加了教育产品的趣味性^[2];在工业方面,交互技术使得机器人具有更强的自动化能力,同时也解放了人的双手,用语音来完成一些控制工作.另外,在娱乐方面,具有交互学习能力的机器人让娱乐产品更加多元化.写字机器人是机器人交互能力的一种重要体现,2017年 CCTV 网络春节晚会上书写春联的机器人 Dobot 让人眼前一亮^[3],受到广泛好评.

历史上,早在 240 年前,瑞士钟表大师 Pierre Jaquet-Droz 已经通过完全机械化的结构,设计了一个会写字的机器人^[4],该机器人内部有 6 000 多个零件,可以书写任何 40 个字母长度的句子和词语,240 年后的今天,这个机器人依然能够精准运行.英文字母只有 26 个,人工的机械化编程可完成所有字母书写,对于比较复杂的汉字来说,如果按照笔画来书写,不同字的书写笔画不同,而且相较于英文单词从左到右的固定结构,汉字的笔画空间结构也极其复杂,因此,按照一定的规则让机械臂实现对汉字正确书写具有一定的编程难度.因此,传统的机械臂写字模型通常都是预先编程,例如写春联机器人 Dobot.但是由于汉字数量庞大,因此完全通过对汉字逐个编程来使机器人完成写字不仅缺乏智能性,也缺乏多样性和交互性.汉字笔画的自动拆分技术为机器人自动写字的可行性提供了基础^[5,6],2015 年第 350 卷《Science》的篇首论文“Human-Level Concept Learning Through Probabilistic Program Induction”^[7],提出了基于笔画部件(component)的计算机自动模仿人类生成文字的方法.然而,这篇论文并没有考虑文字的笔顺问题.如何让机器人能够与用户交互,并且通过交互学习汉字书写方式,包括笔画与笔顺,是机器人写字具有交互友好型的重要内容.

为了提升写字机器人的智能性,使得机器人能够与人进行交互并学习新的汉字书写方式,本文基于 Uarm 机械臂构建了一个可以写字的机械臂写字系统,该系统融合语音、图像等交互技术,不仅可以对看到的字进行自动笔画拆分与书写,而且可以通过与用户的交互学习正确的汉字书写过程,该系统通过智能手段为用户提供了交互书写的乐趣,在儿童教育方面具有重要价值.

本文第 2 节首先介绍本文系统的整体架构.第 3 节对系统中的主要环节进行介绍,并详细介绍交互流程.第 4 节为实验及分析部分.第 5 节为总结.

2 系统框架

本文系统为一个基于 Uarm 机械臂的写字系统,本节主要介绍该系统的整体架构,对系统的整体结构从宏观上进行阐述,系统中的一些关键技术内容在下一节详细描述.本文系统的整体结构如图 1 所示.

系统共分为 3 个主要模块:信息输入、关键技术以及输出反馈.系统的输入信息模块包含用户的语音信息以及摄像头观察到的文字的图像信息.图 1 所示灰色部分为本系统关键技术模块,主要包含两大部分,一是通过对用户的语音信息进行分析,可以获得用户想要写的关键字及用户意图,并根据当前状态进行对话管理;二是通过对摄像头看到的图像信息进行分析,对检测到的汉字进行自动笔画拆分和笔顺提取,对于正在教授的字,跟踪

笔迹顺序,学习新写法.最后一个模块是输出反馈,通过对话管理,机械臂会以对话的形式进行反馈,与用户交互,并根据调用机械臂的写字程序,书写需要写的字.下面对每一个功能进行详细说明.

语音信息:为用户与机械臂进行交互的方式,用户通过麦克风与机械臂系统以对话的形式来完成交互,使得机械臂系统完成交互任务.

图像信息:摄像头会实时检测场景中是否有要写的字,如果有,会记录该字并自动提取笔画,另外,在用户教机械臂写字的过程中,摄像头会实时记录用户教的笔画顺序,并进行学习和记忆.

对话管理:管理对话内容,控制整个系统的对话流程,也包括对话中的关键字提取,用户在让机械臂写字时,需要明确告诉机械臂要写哪个字,比如“写一个天气的天字”,系统会提取到要写的关键字为“天”,然后去查看是否会写.

意图分类:将用户的对话意图进行分类,本系统共设置 Write、Teach、Positive、Negative、Others 等意图,其中, Write 表示让机械臂写字,Teach 表示教机械臂写字,Positive 表示肯定或正面评价,Negative 表示否定或负面评价,Others 表示其他意图,是闲聊等与书写无关的意图.

自动笔画拆分:机械臂对看到的字提取特征点,然后进行自动笔画拆分,并基于规则进行笔画顺序规划.

笔顺提取:对于用户正在教授的字,提取用户笔画顺序,学习记忆新的写字方式.

语音输出:通过对话管理,将需要反馈的信息以语音形式输出,与用户进行交互.

机械臂写字:根据自动拆分笔画后提取的笔画顺序或者用户教授的笔画顺序写字.

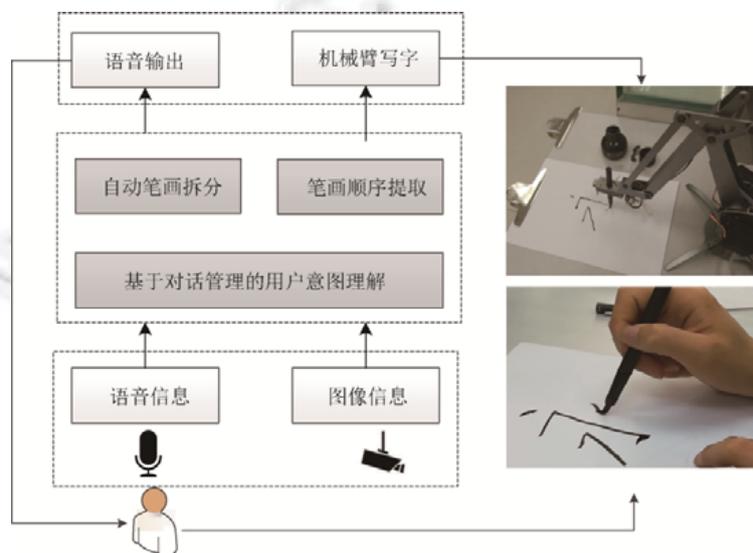


图 1 智能交互机械臂写字系统框架

3 智能交互写字机械臂

本文系统中用到的主要关键技术为笔画拆分、笔画顺序提取以及对话管理,下面对这 3 部分进行详细介绍.

3.1 对话管理

对话管理是对系统状态进行跟踪,对整个交互流程进行管理^[8,9].对话管理的具体流程如图 2 所示.

图中 S_0, S_1, \dots, S_4 分别表示不同阶段系统所处的状态,具体每个状态的描述见表 1.图 2(a)所示为系统从 S_0 初始状态开始,对于要书写的字每一步的对话流程,图 2(b)所示为系统从 S_3 开始进行的对话管理流程,其中, S_3 状态为进入用户教授机械臂写字的状态.

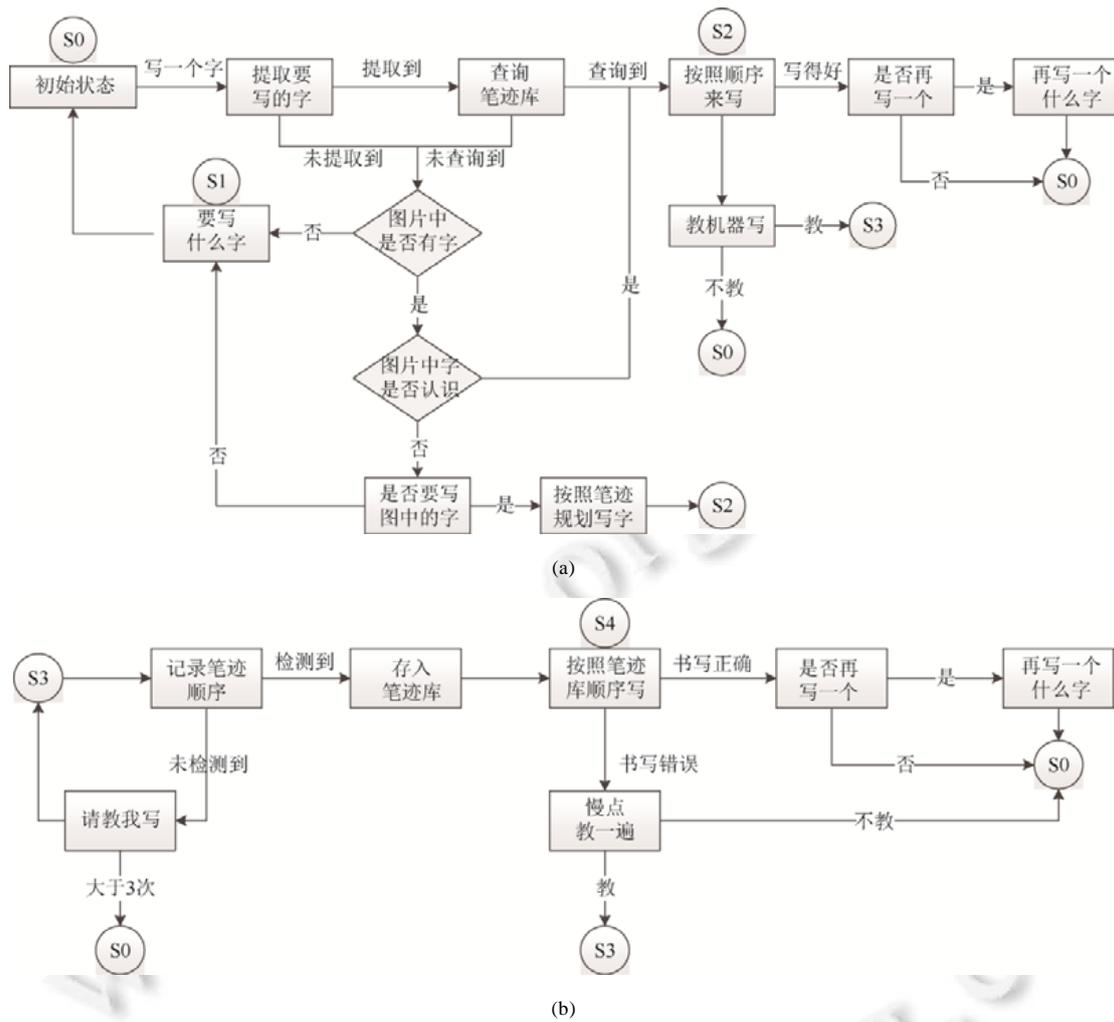


图 2 对话管理具体流程

表 1 对话管理中各状态描述

状态	状态描述
S0	初始状态
S1	提取到关键字,但不会写,询问是否可以教
S2	写完字,请求评价
S3	询问是否可以教机械臂写字
S4	教机械臂写字完成,开始学写一遍

3.2 基于特征点提取的笔画拆分

该部分的主要功能是对一个陌生的字实现基于简单规则的笔画规划,主要由 3 部分组成:汉字骨架抽取、特征点提取以及汉字笔画规划.抽取汉字骨架的目的是为了更准确、更方便地对汉字的特征点进行抽取,进而对笔画进行规划.

为了达到更好的实验效果,本文首先对待处理的汉字图像进行了二值化处理.由于实验光照环境的不确定,二值化处理最佳阈值往往是随环境变化而变化的,为解决此问题,本文采用最大类间方差法^[10,11]来确定最优阈值.对汉字图像进行二值化处理后,再对该二值图像进行遍历,遍历所有边缘点,并且删除可以被删除的边缘点,

然后重复上面步骤,直到没有可被删除的点则说明已完成了对汉字图像骨架的抽取.

本文使用 Harris 角点检测^[12,13]算法来提取汉字骨架上的特征点,该算法首先设定一个核窗口,然后用此窗口在图像上遍历所有像素点,如果某像素点和周围任意方向的核窗口区域灰度值之和相差都较大,则说明该像素点的角点强度较大,以此来寻找汉字骨架中的特征点.实验效果图如图 3 所示.

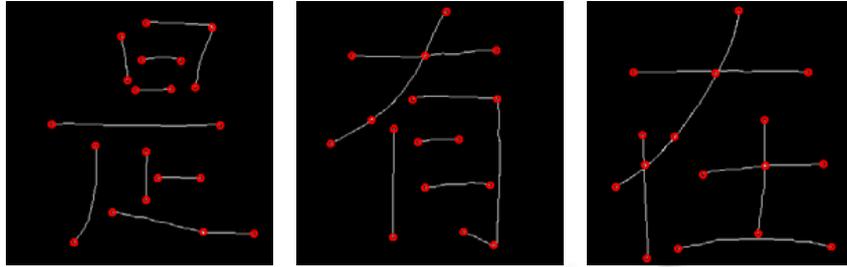


图 3 汉字特征点提取效果图

提取汉字特征点之后,需要对汉字的笔画进行规划,本文主要采用了几个简单的规则,并不能完美地归纳出正常人正常写字笔顺,故有时基于这些简单规则的笔画规划顺序出现错误在所难免.本文笔画规划的主要流程是,首先确定下笔点,下笔点是汉字最左上角的一点,然后遵循从左到右从上到下的习惯顺序规划笔画.整个笔画规划流程如图 4 所示.

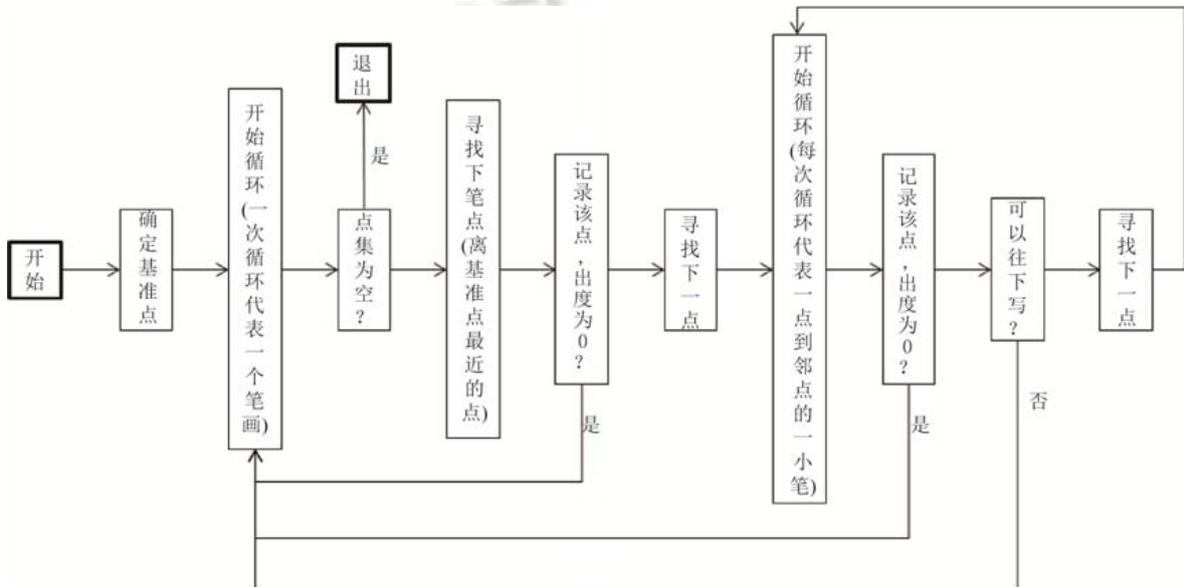


图 4 笔画规划总体流程图

3.3 笔画顺序提取

笔画顺序提取的主要功能是对书写汉字的笔顺加以提取,主要有 4 个步骤:图像预处理、汉字骨架提取、笔画聚类以及特征点检测,如图 5 所示.

系统在通过摄像头记录书写汉字每个像素点的成像时间时,由于在原始图像采集的过程中受到笔的落笔位置以及采集过程中光照变化的影响,会给汉字成像的时间记录带来一定程度的困难,为了更好地对汉字笔画的时间顺序进行记录,首先对原始图像进行预处理,采用最大类间方法进行阈值分割,提取汉字笔画的二值图像,同时对图像进行去噪;然后对预处理的二值图像进行骨架提取,通过图像的二值化处理和汉字的骨架提取能

够很好地去除影响记录汉字笔画的时间顺序,获取汉字骨架图像中每个像素点的记录时间。

通过对汉字骨架中的时间顺序进行聚类,根据人们手写汉字的习惯,将间隔时间超过一定阈值的时间顺序进行聚类,获取汉字的每一笔画,然后通过上文所描述的使用 Harris 角点检测算法来提取汉字骨架上每一笔画的特征点,对比每一笔画特征点的时间记录,获取笔画中特征点的时间顺序,以此来进行汉字笔顺的提取。

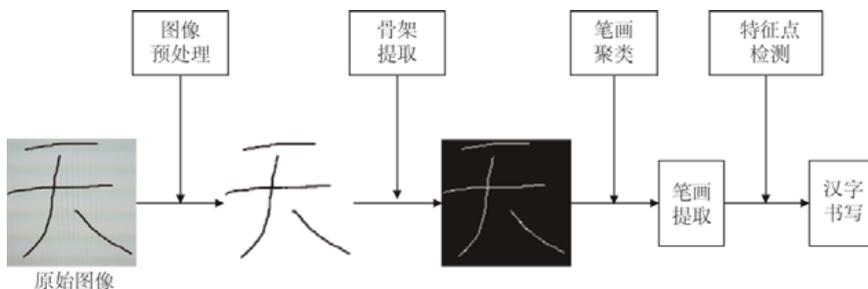


图 5 汉字笔画顺序提取

4 实验分析

本节对系统进行实验和分析,以验证该系统的写字效果及用户对该系统的主观评价。

4.1 对话意图

对整个系统而言,交互的体验感十分重要.而交互的体验主要体现在对话过程中,由于语言识别或者语义理解的错误,有时会对用户对话意图的判断造成一定的影响.本实验对 150 轮中的 634 句有效对话的用户意图识别进行测试,并分析在语音识别的准确率对对话意图分类的影响.实验结果见表 2.

表 2 用户对对话意图识别准确率

	语音识别正确	语音识别正确情况下的对话意图	语音识别错误情况下的对话意图
正确对话数	511	443	62
准确率	0.819	0.867	0.549

从实验结果可以看出,语音识别的正确率在 81.9%,在语音识别正确的情况下,对话意图的正确率为 86.7%,而在语音识别错误的情况下,对话意图准确率仅为 54.9%,由于语音识别的错误,导致对话意图准确率下降了 32%左右.由此可见,语音识别的准确性对对话意图的影响很大,而且也直接影响到交互的效果。

4.2 拆字准确率分析

对于摄像头看到的字,系统会自动提取该字的关键点,并对该字以笔画为基础进行拆分,然后对拆分的结果进行笔画顺序规划.本实验针对各种字体 587 个字进行拆分,其中包含宋体、楷体标准字体及不同人的手写字体.实验结果见表 3.

表 3 拆字实验准确率

	特征点提取	笔画拆分	笔画顺序
正确字个数	534	501	56
准确率	0.910	0.853	0.095

通过实验可以发现,对于特征点的提取,具有 91%的准确率,相对于系统而言,比较可靠,而且笔画拆分的正确率达到 85.3%,也基本满足系统要求,但对于笔画顺序的规划,正确率却很低,仅为 9.5%,很大的原因在于汉字的空间结构复杂,如果仅按照人类书写的笔画规则,对计算机而言,还是很难掌握正确的笔画顺序,因此提升机械臂写字的笔画顺序对系统来说尤为关键。

4.3 笔画顺序准确率分析

由第 4.2 节给出的实验可知,系统对汉字的自动笔画拆分正确率基本满足系统需求,但笔画顺序的正确率却很低,因此需要对机械臂写错的字进行人工教授,教其学会正确的笔画顺序.本实验主要针对用户教授的笔画顺序进行,以测试机械臂在学习之后的学习情况.对于重复写错的字,最多只能教 3 次,本实验对 67 名用户教授的 358 个有效结果进行评测,实验结果见表 4.

表 4 笔画顺序提取准确率

	特征点提取	笔画顺序
正确字个数	330	319
准确率	0.922	0.891

通过实验可以发现,对于特征点的提取,准确率为 92.2%,与之前的 91% 准确率相比,基本不变,但笔画顺序的正确率却达到了 89.1%,远远超过自动拆分后笔画规划的结果,可以看出,系统在学习用户教授写字上,取得了不错的效果.

4.4 主观结果评测

由于本文系统是一个交互系统,用户的交互体验对系统来说也是很重要的.因此,我们邀请 67 位参与者参与机械臂系统的体验和评测,每人至少进行 3 轮以上的有效对话,最后在其他评测结束后,要求每个体验者对系统进行满意度投票,总共 5 个选项,分别是很满意、满意、一般、不太满意和很不满意,其统计分布如图 6 所示.

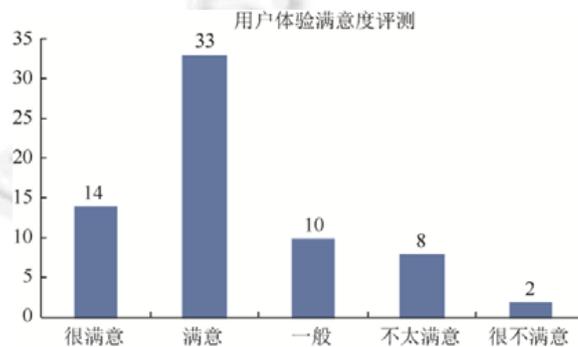


图 6 用户体验满意度评测

由图 6 评测结果可知,70.1% 的人对交互体验感觉满意或者很满意,而只有 14.9% 人对系统体验不太满意或者很不满意.从用户的主观评测可以看出,该系统的交互体验非常不错,能够获得大多数体验者的认可.

5 总结

本文基于 Uarm 机械臂构建了一个学习人类写字顺序的机械臂智能写字系统,该系统首先具有对陌生汉字的自动笔画拆分和书写能力,然后基于语音对话和图像分析技术,能够根据用户教授的笔画和笔顺来学习汉字的正确书写方式.首先,系统根据输入的语音信息以及摄像头观察到的文字的图像信息,获得用户想要写的关键字及对话意图;然后通过对摄像头看到的图像信息进行分析,对检测到的汉字进行自动笔画拆分和笔顺提取,对于正在教授的字,跟踪笔迹顺序,学习汉字笔顺的正确写法.通过对话管理,机械臂会以对话的形式进行书写反馈并与用户交互,学习人类书写顺序并实现正确书写.通过实验分析及测试者的主观评测,该系统取得了不错的评价.目前系统中仍然存在一些问题,如系统目前仅能通过观测得到的用户笔顺去书写已知汉字.人类具有从已知汉字的学习过程中归纳出陌生汉字正确书写顺序的能力,在这一点上,如何使得机械臂具有自主学习并从历史经验中归纳出陌生汉字的书写能力,是提升机械臂智能的关键,也是我们未来的一个研究工作.

References:

- [1] Rogers Y, Sharp H, Preece J. Interaction design: Beyond human-computer interaction. *Journal of Neuroscience the Official Journal of the Society for Neuroscience*, 2012,2002(4):369–378.
- [2] Chin KY, Hong ZW, Chen YL. Impact of using an educational robot-based learning system on students' motivation in elementary education. *IEEE Trans. on Learning Technologies*, 2014,7(4):333–345.
- [3] <http://ent.sina.com.cn/2017-01-19/doc-ifxzuswr9412496.shtml>
- [4] <http://mt.sohu.com/20150609/n414676188.shtml>
- [5] Su YM, Wang JF. A novel stroke extraction method for Chinese characters using Gabor filters. *Pattern Recognition*, 2003, 36(3):635–647.
- [6] Liu CL, Kim IJ, Jin HK. Model-Based stroke extraction and matching for handwritten Chinese character recognition. *Pattern Recognition*, 2001,34(12):2339–2352.
- [7] Lake BM, Salakhutdinov R, Tenenbaum JB. Human-Level concept learning through probabilistic program induction. *Science*, 2015,350(6266):1332.
- [8] Mctear MF. Modelling spoken dialogues with state transition diagrams: Experiences with the CSLU toolkit. In: *Proc. of the Int'l Conf. on Spoken Language Processing*. 1999. 1223–1226.
- [9] Lee C, Cha YS, Kuc TY. Implementation of dialogue system for intelligent service robots. In: *Proc. of the Int'l Conf. on Control, Automation and Systems*. IEEE, 2008. 2038–2042.
- [10] Tanabe N, Yamamoto J, Nishi H, *et al.* MEMOnet: Network interface plugged into a memory slot. In: *Proc. of the IEEE Int'l Conf. on Cluster Computing*. IEEE, 2000. 17–26.
- [11] 肖超云,朱伟兴.基于 Otsu 准则及图像熵的阈值分割算法. *计算机工程*,2007,33(14):188–189.
- [12] Smith SM, Brady JM. SUSAN—A New Approach to Low Level Image Processing. Kluwer Academic Publishers, 1997.
- [13] 赵万金,龚声蓉,刘纯平等.一种自适应的 Harris 角点检测算法. *计算机工程*,2008,34(10):212–214.



杨明浩(1977—),男,四川达州人,博士,副研究员,CCF 专业会员,主要研究领域为人机融合感知与决策,交互学习认知计算理论与方法,多通道交互信息处理.



张珂(1991—),男,硕士,主要研究领域为人机口语对话.



赵博程(1989—),男,硕士,主要研究领域为认知计算.



朱庆杰(1988—),男,硕士,主要研究领域为图像理,手势跟踪.



潘航(1991—),男,硕士,主要研究领域为图像处理.



那燊若阳(1994—),男,硕士,主要研究领域为图像处理技术.



湛永松(1979—),男,博士,CCF 专业会员,主要研究领域为数字媒体技术,计算机图形图像技术.



陶建华(1971—),男,博士,研究员,博士生导师,CCF 会士,主要研究领域为语音合成,语音识别,情感计算,人机对话.